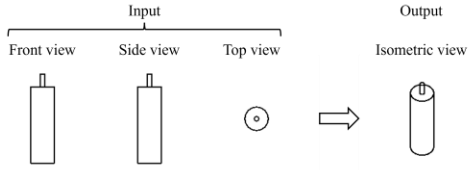
Isometric View Images Generation from Three Orthographic View Contour Drawings using Enhanced IsoGAN

Thao Nguyen Phuong[1,2], Hidetomo Sakaino[1], Vinh Nguyen Duy[1] {ThaoNP47, HidetomoS, VinhND1}@fpt.com, nguyenphuongthao@akane.waseda.jp
[1]AI-Image Group, Data Solution Dept., (FCJ.ABC) FPT Consulting Japan, FPT Software
[2]Department of Computer Science and Communications Engineering, Waseda University

FPT Japan Holdings WASEDA University

CVPR JUNE 17-21, 2024 SEATTLE, WA

Three view images of an object and its corresponding isometric view images



Motivation

- CAD software plays a pivotal role in generating 3D shapes across various industrial domains.
- Orthographic projection is the most popular method used to depict 3D objects onto a 2D surface. 2D orthographic view drawings are usually utilized by designers to present their ideas.
- Isometric view images are very important for 3D reconstruction process because they keep much information of the 3D object.

⇒ A method capable of transforming three-view contour drawings into isometric view images is highly desired.

- Existing works deal with the task as a sequence-to-sequence problem, producing sequences of CAD commands for 2D drawing or 3D object reconstruction as output instead of images.

Statement

- Isometric image generation from three orthographic views contour drawings can be cast as **multi-images-to-image translation** problem.
- Unlike common image-to-image translation tasks, there is a strong spatial and geometrical relation between each orthographic view of the object.
- SOTA generative models only take one image or text as input and map it to the target domain without considering the relation between multiple input images.

Contributions

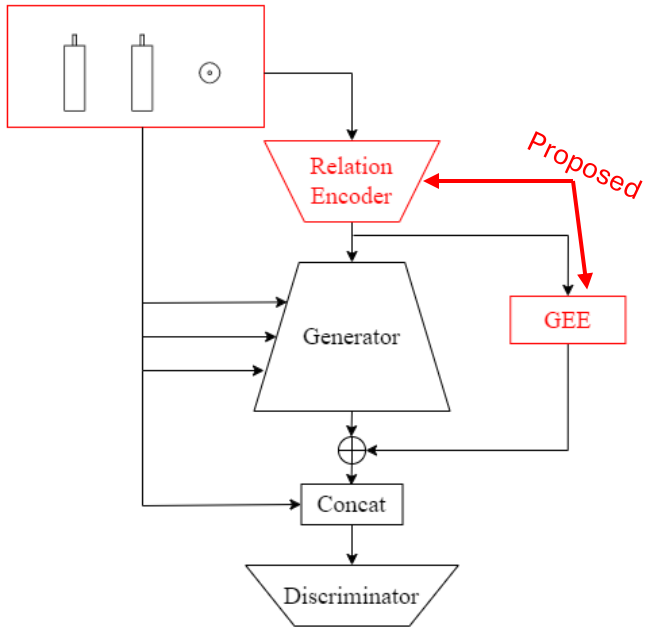
- The spatial and geometrical relation between each orthographic view contour drawing of an object on raster image is investigated using an **encoder network**.
- We propose **IsoGAN**, a GAN-based model with a novel **Gaussian Enhanced Euclidean norm (GEE)** attention block. A **modification of the GAN loss function** is also presented.

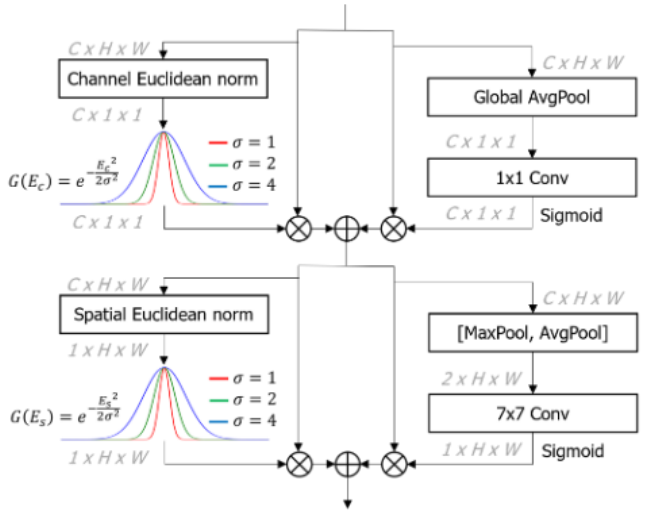
Conclusion

This paper has proposed a novel IsoGAN framework for efficiently automatic isometric view image generation from three orthographic views contour drawings. An encoder is deployed to analyze the spatial and geometrical relation between each view of the object and transform these relations into a vector, which then is taken as an extra input to the generator. An attention mechanism and modified loss function are also proposed to refine the generation result.

Proposed method

Three-views images of an object





The overall architecture of the proposed IsoGAN (left). Gaussian Enhanced Euclidean norm (GEE) attention block details (right). ⊗ represents broadcast element-wise multiplication and ⊕ denotes element-wise addition. Ec and Es represent Channel Euclidean norm and Spatial Euclidean norm, respectively. A triple of three-view contour drawings is input to the encoder, and is consecutively upsampled to feed to each stage of the generator.

Three-view contour drawings encoder

An encoder is employed to transform a triplet of front, side, and top view image into a vector, subsequently provided to the generator. Loss function:

$$\mathcal{L}_{\mathrm{KLD}} = \mathcal{D}_{\mathrm{KL}}(q(\mathbf{z} \mid \mathbf{x}) \| p(\mathbf{z})),$$

Isometric view image generation

IsoGAN is an adaption of SPADE and MoNCE with a novel attention mechanism Gaussian Enhanced Euclidean norm, which applies Gaussian function to Channel and Spatial Euclidean norm.

$$\mathbf{F}' = \mathbf{M_c}(\mathbf{F}) \otimes \mathbf{F} \oplus \mathbf{G}(\mathbf{E_c}) \otimes \mathbf{F}$$
$$\mathbf{F}'' = \mathbf{M_s}(\mathbf{F}') \otimes \mathbf{F}' \oplus \mathbf{G}(\mathbf{E_s}) \otimes \mathbf{F}',$$

Modification to loss function: $\mathcal{L} = GANLoss + MoNCE + \ell(\hat{p}_i, p_i)$,

References

[1] Han, Wenyu, Siyuan Xiang, Chenhui Liu, RuoyuWang, andChen Feng. "Spare3d: A dataset for spatial reasoning on three-view line drawings." In Proceedings of the IEEE/CVF Conference onComputer Vision and Pattern Recognition, pp. 14690-14699. 2020. [2] Park, Taesung, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. "Semantic image synthesiswith spatially-adaptive normalization." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2337-2346. 2019. [3] Zhan, Fangneng, Jiahui Zhang, Yingchen Yu, Rongliang Wu, and Shijian Lu. "Modulated contrast for versatile image synthesis." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18280-18290. 2022. Ruan, Dongsheng, Daiyin Wang, Yuan Zheng, Nenggan Zheng, and Min Zheng. "Gaussian context transformer." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15129-15138. 2021. [4] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. "Cbam: Convolutional block attention module." In Proceedings of the European conference on computer vision (ECCV), pp. 3-19. 2018.

Dataset

SPARE3D was used to conduct our experiments, contains 5000 pairs of three-view and isometric image.
The number of training and testing pairs is 4000 and 1000, respectively.

Experiment Setup

NVIDIA RTX 4070 GPU with 12GB VRAM, PyTorch framework.
Frechet Inception Distance (FID), Structural Similarity Index (SSIM), L1, and L2 metrics are used in our experiments.

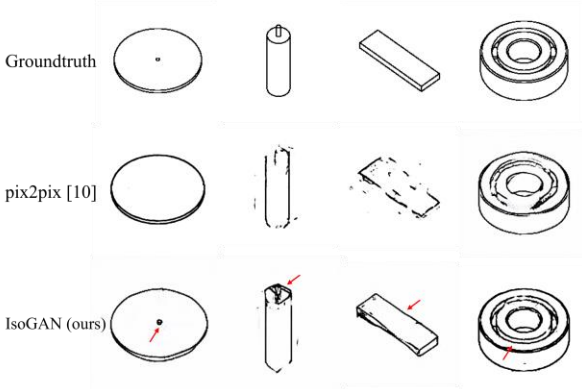
Experimental results

Quantitative Results

Table 1. Quantitative generation performance comparisons with baseline model.

Method	FID	SSIM	L1	L2
pix2pix (baseline)	43.82	0.436	23.34	21.73
IsoGAN (ours)	**22.59**	**0.651**	**16.85**	**15.29**

Qualitative Results



Qualitative comparisons of isometric view generation performance with baseline model.

Our proposed method generates more similar isometric view images compared to groundtruth images. Red arrows denote where IsoGAN reconstructs better contours compared to the baseline approach.